UDC 551.577.21:551.547.5:551.511.12(73)

# CORRELATIONS BETWEEN AREAL PRECIPITATION AND 850-MILLIBAR GEOPOTENTIAL HEIGHT

IVER A. LUND

Air Force Cambridge Research Laboratories, Bedford, Mass.

## ABSTRACT

Geopotential heights of the 850-mb surface at 499 grid points over one-half of the Northern Hemisphere, extending from 10°E westward to 170°W, were correlated with daily precipitation observations over California and an area in the Eastern United States on 401 January and February days. Correlations as high as 0.69 were obtained between 850-mb heights and California precipitation. The statistical significance of the maximum correlation obtained over the one-half hemisphere area exceeded the 5-percent level for all periods of less than 5 days. The maximum correlation between heights and precipitation observed in the Eastern United States was −0.44 and the correlations were significant at the 5-percent level for all periods of less than 3 days.

Since correlations between 850-mb heights and subsequent precipitation exceed autocorrelations except for short periods of time, it is proposed that 850-mb height observations replace or supplement persistence in conditional climatologies of precipitation.

## 1. INTRODUCTION

It is generally assumed that the probability of future precipitation can be estimated from a long record of historical data, that a sharper estimate is possible if the current weather is considered, and that the probability estimates can be further sharpened by including information about the current atmospheric circulation. Useful models have been developed by Gringorten (1966) to estimate the duration of weather events and McAllister (1969) to estimate conditional probabilities of recurrence of weather events, but the problem of extending such models to include circulation parameters is unsolved. Probabilities of weather events associated with map types such as those developed by Krick (1943) and Elliott (1951) are useful aids to prediction but maps are often difficult to type objectively, and difficult to arrange in order of increasing probability of a given weather event. Maps can be objectively typed by the method of Lund (1963) but arrangement in a meaningful order with respect to a given weather event is difficult.

Klein et al. (1968) have shown that winter precipitation in the western plateau states is more closely related to the circulation at the 850-mb surface than at any other upper level. Other studies by Klein (1968) and Russo et al. (1966) show that the lower the level, the better appears to be the relationship to precipitation.

Because a more complete record of 850-mb data was available than for lower levels, 850-mb data were used in this study to identify predictors which might be used to replace or supplement persistence in conditional climatologies of precipitation. The study also includes information on the time limit of predictability of areal precipitation when a single 850-mb height is the only predictor.

## 2. DATA

The daily total precipitation observed between 0000 and 2400 LST, at 10 stations in each of two areas, was used as the predictand. The stations are listed below and the areas are shown in figure 1.

| | *Eastern United States* | *California* |
|---|---|---|
| 1. | Concord, N.H. | Red Bluff |
| 2. | Boston, Mass. | Chico |
| 3. | Hartford, Conn. | Colusa |
| 4. | Albany, N.Y. | Brooks |
| 5. | New York, N.Y. | Sacramento |
| 6. | Harrisburg, Pa. | Stockton |
| 7. | Philadelphia, Pa. | Modesto |
| 8. | Atlantic City, N.J. | Merced |
| 9. | Baltimore, Md. | Fresno |
| 10. | Washington, D.C. | Angiola (Corcoran) |

The data include all January and February days from 1962 through 1969.

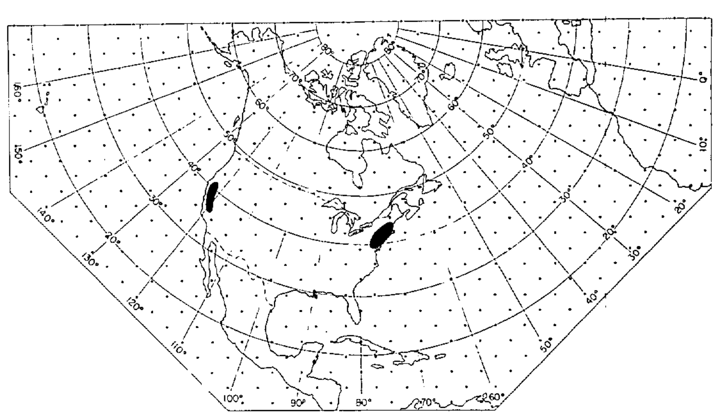

FIGURE 1.—Location of the two precipitation areas and the 499 grid points over which geopotential heights were obtained.

Heights of the 850-mb surface at the 499 grid points shown in figure 1 were also studied. Data were available for 401 of the possible 424 days between January 1 and February 22 for the years from 1962 through 1969.

TABLE 1.—*Ten-station correlations between combined daily precipitation and that 1–6 days later; Jan. 1–Feb. 22, 1962–1969*

| Days after precip. | Eastern U.S. | Calif. |
|---|---|---|
| 1 | 0.01 | 0.63 |
| 2 | −0.09 | 0.20 |
| 3 | −0.03 | 0.16 |
| 4 | −0.04 | 0.12 |
| 5 | −0.02 | 0.11 |
| 6 | −0.06 | 0.14 |

## 3. CORRELATIONS

The total precipitation observed at the 10 stations in each area during the period from January 1 through February 22 for the years 1962 through 1969, a sample of 424 days, was correlated with the precipitation observed from 1 to 6 days later (table 1). The correlation coefficients provide a measure of precipitation persistence.

Geopotential heights were correlated with precipitation using the correlation coefficient

$$r=\frac{N\Sigma ZY-\Sigma Z\Sigma Y}{\{[N\Sigma Z^2-(\Sigma Z)^2][N\Sigma Y^2-(\Sigma Y)^2]\}^{1/2}} \tag{1}$$

where $N$ equals 401, $Y$ is a precipitation amount, and $Z$ is a height value.

About 7,000 correlation coefficients were computed to prepare 14 maps depicting relationships between height

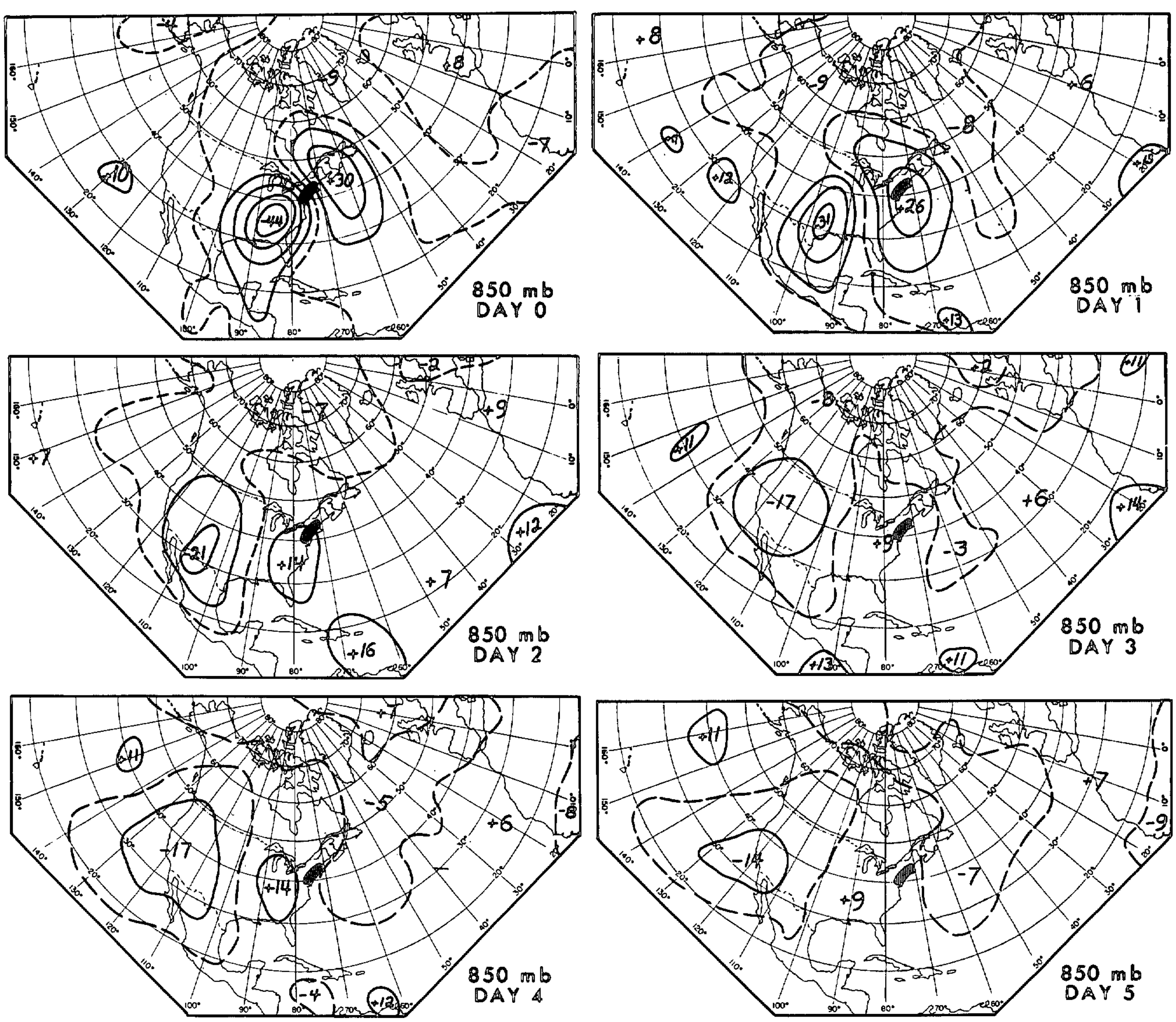


FIGURE 2.—Correlations between 0700 LST 850-mb geopotential heights and total precipitation observed between 0000 and 2400 LST at 10 stations in Eastern United States (darkened area), on 401 January and February days, for lags of 0 through 5 days.

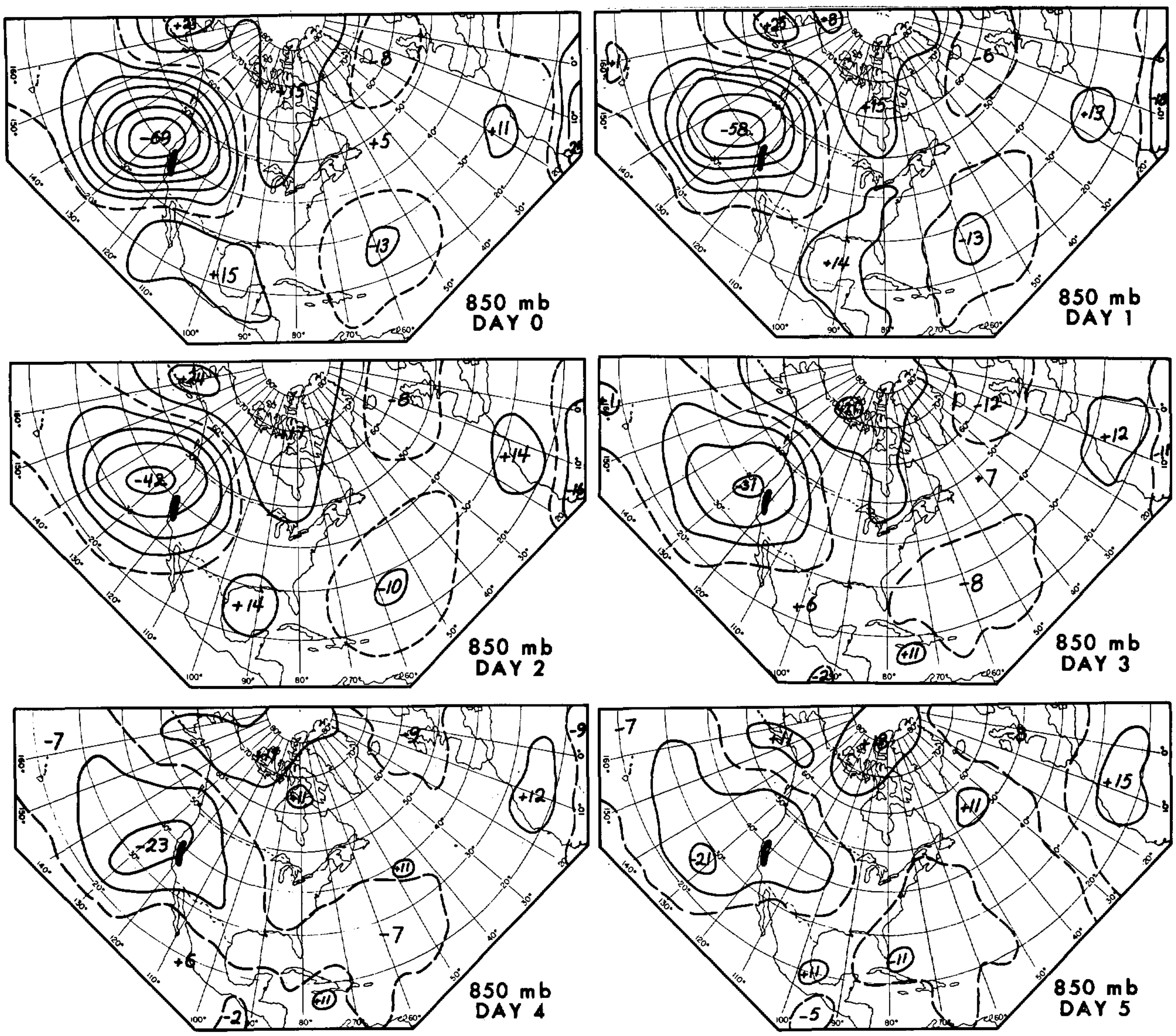


FIGURE 3.—Correlations between 0400 LST 850-mb geopotential heights and total precipitation observed between 0000 and 2400 LST at 10 stations in California (darkened area), on 401 January and February days, for lags of 0 through 5 days.

data and precipitation 0, 1, 2, 3, 4, 5, and 6 days later at each of the two areas. Maps depicting correlations between 850-mb heights and Eastern United States precipitation are shown in figure 2. Correlations with California precipitation are shown in figure 3. The maps for the 6-day lag are omitted to conserve space.

The tracks of the centers of maximum correlation shown in figures 2 and 3 are depicted in figure 4. There is good continuity from day to day, indicative of a genuine physical relationship between 850-mb geopotential heights and subsequent areal precipitation.

## 4. SIGNIFICANCE OF EASTERN UNITED STATES CORRELATIONS

Standard statistical tests do not apply to the correlation coefficients computed in this study for the following reasons: (1) precipitation values used in this study are not normally distributed in a statistical sense, (2) geopotential height data are correlated both serially and spacially, and (3) some serial correlation is present in California precipitation data. Therefore, the method proposed by Lund (1970) was used to test the statistical significance of the coefficients.

The frequency distribution of the 424 daily Eastern United States precipitation values is shown in figure 5. The values greater than zero were fitted to a normal distribution. The resulting curve

$$\hat{Y}=0.103+0.173X+1.237X^3+0.171X^4-0.251X^5+0.041X^6 \quad \text{when } X\geq -0.3 \tag{2}$$

and

$$\hat{Y}=0 \quad \text{when } X<-0.3,$$

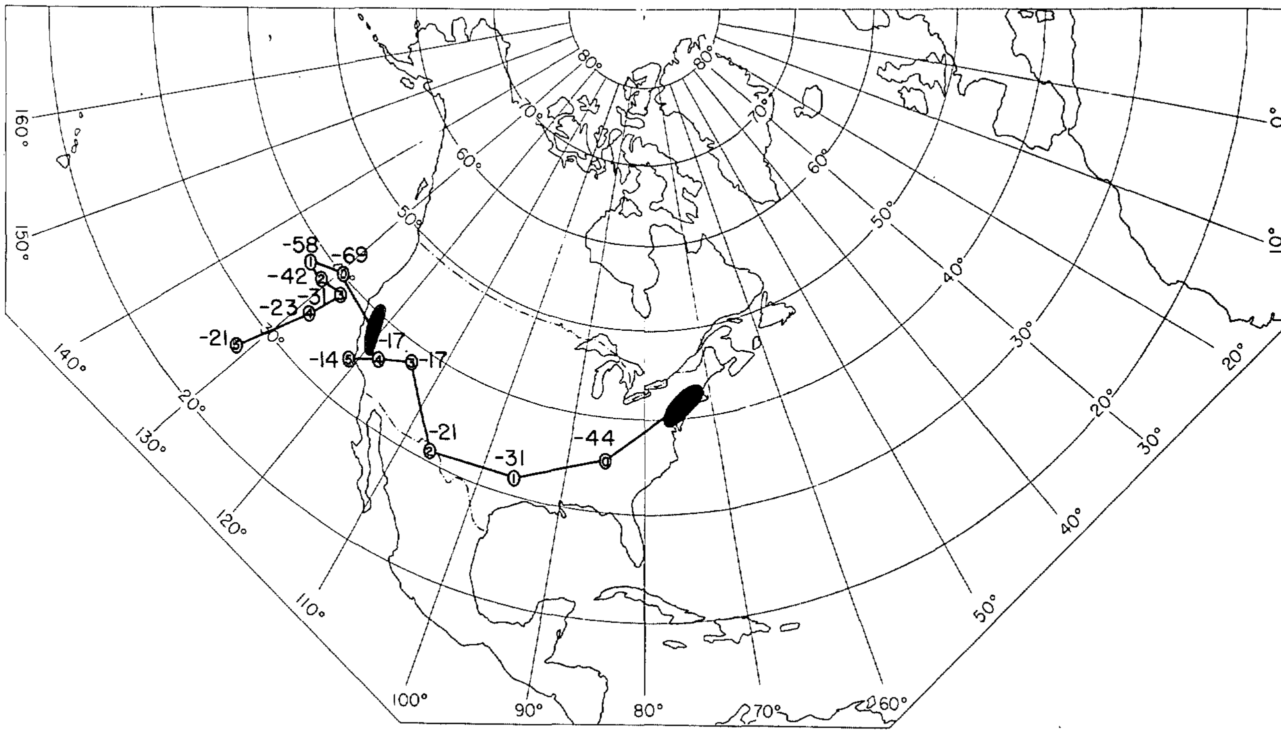


FIGURE 4.—Locations of centers of maximum correlation between areal precipitation (darkened areas) and 850-mb geopotential heights for lags of 0 through 5 days.

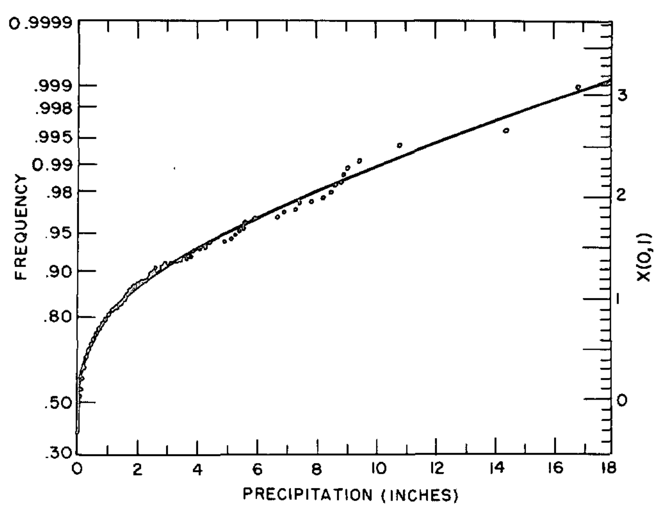


FIGURE 5.—Frequency distribution of the 424 daily 10-station total January 1–February 22 precipitation values for Eastern United States (1962–1969). The curve is a least-squares fit to normalized values, eq (2).

where $\hat{Y}$ is an estimate of Eastern United States precipitation and $X$ is a normal number, is also shown in figure 5.

Equation (2) was used to generate 8,020 bogus precipitation values from 8,020 random normal numbers. The 8,020 values were divided into 20 sets of 401 values. Each set of 401 values was correlated with the set of 401 heights at each of the 499 grid points. The largest of the 499 coefficients obtained with each of the 20 sets of bogus precipitation values was −0.137, +0.122, +0.118, +0.172, −0.146, +0.128, −0.170, +0.137, +0.135, −0.119, +0.143, +0.111, −0.113, −0.142, −0.158, −0.169, −0.167, +0.103, +0.146, and −0.120 for computer runs 1 through 20, respectively. These values, without regard to sign, are plotted in figure 6 using the rule of Blom (1958)

$$\hat{P}_i=(i-3/8)/(n+1/4) \qquad (3)$$

where $i$ is the order of the observation and $n$ is the number of observations. A smooth curve was drawn through the points.

The highest correlations, in absolute values, obtained on corresponding maps when actually observed Eastern United States precipitation data were used, are also plotted in figure 6. Correlations exceeding the 0.0001 level of significance were found for periods up to at least 1 day. The chosen significance level of 0.05 was exceeded for all lags of less than 3 days.

Since the test used for estimating the statistical significance of the highest computed correlation coefficient in a set of intercorrelated correlation coefficients is not widely accepted, the distribution of all 9,980 correlation coefficients computed in the 20 runs, with bogus precipitation data, was determined and plotted in figure 7.

Also shown in the figure is the expected frequency distribution in the particular case $\rho=0$ and $N=401$ days, where $\rho$ is the population coefficient. Since the 850-mb geopotential heights ($Z$s) are approximately normally distributed, the fact that the $Y$s (Eastern United States precipitation) are far from normally distributed does not change the distribution of the sample correlation coefficient, $r$. This fact has been proved by Fisher (Snedecor and Cochran 1967). Since the distribution is symmetrical,

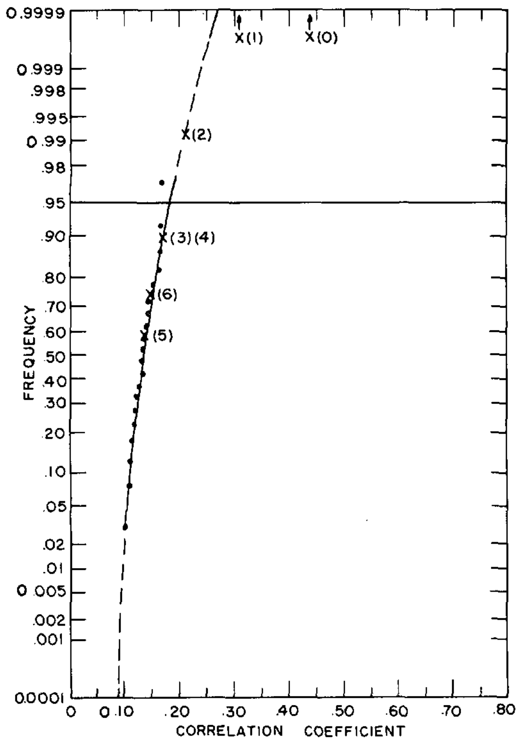


FIGURE 6.—Frequency distribution of the highest (in absolute value) of 499 correlations on each of 20 charts of correlations between 850-mb heights and bogus precipitation values (dots), and corresponding correlations between the same 850-mb heights and actually observed Eastern United States precipitation ($X$s) on the same day (0), and 1 through 6 days after the heights were observed.

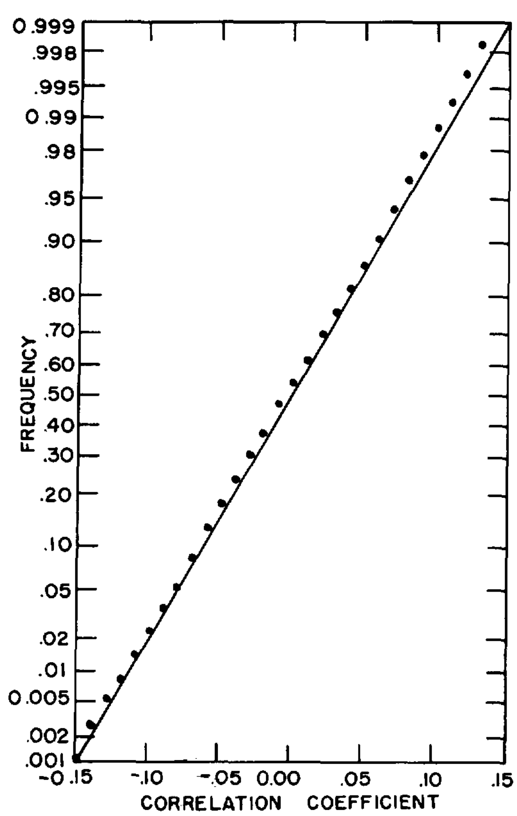


FIGURE 7.—Frequency distribution of 9,980 correlation coefficients computed between 850-mb geopotential heights observed on 401 days at 499 grid points and for 20 sets of bogus Eastern United States precipitation values. The line is the expected distribution when $\rho=0$.

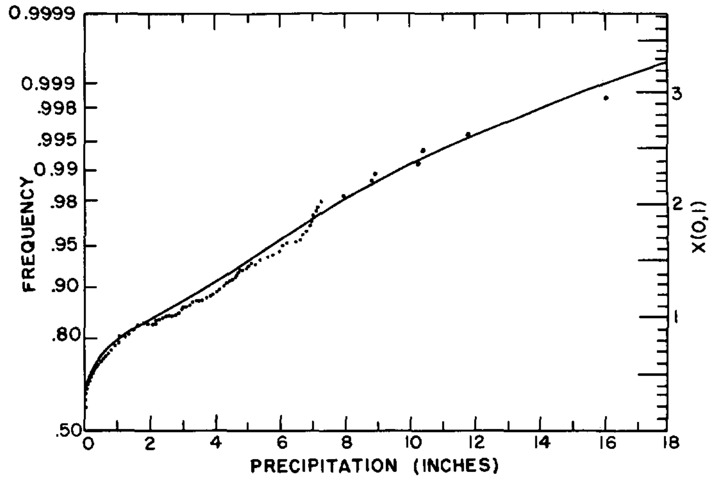


FIGURE 8.—Frequency distribution of the 424 daily 10-station total January 1–February 22 precipitation values for California (1962–1969). The curve is a least-squares fit to normalized values, eq (4).

the sign of $r$ is ignored when testing the significance of the $r$ values computed between pairs of actually observed data.

## 5. SIGNIFICANCE OF CALIFORNIA CORRELATIONS

The frequency distribution of the 424 daily California precipitation values and a least-squares curve fit to the data,

$$\hat{Y}'=0.301X_1-4.378X_1^2+12.817X_1^3-9.065X_1^4+2.656X_1^5-0.279X_1^6 \quad \text{when } X_1\geq 0.4,$$

and (4)

$$\hat{Y}'=0 \quad \text{when } X_1<0.4,$$

where $\hat{Y}'$ is an estimate of California precipitation and $X_1$ is a normal number, is shown in figure 8. The California rainfall distribution is very similar to the Eastern United States distribution except that days with little or no rainfall, less than 0.06 in. for example, are more common in California (59 percent) than in the East (49 percent).

Bogus precipitation values were generated from a new set of 8,020 random normal numbers in the same manner as described for the Eastern United States significance test. The highest correlation coefficients obtained from the 20 sets of 499 coefficients each, when bogus precipitation values were correlated with 850-mb heights, are plotted in figure 9. Also plotted are the correlations obtained when actually observed California precipitation values were used. Correlations based on actual data are significantly higher than expected by chance for all periods of less than 5 days.

Because serial correlation has the effect of reducing the sample of independent cases, an equivalent correlation was introduced into the random normal numbers before the significance test was applied to make the bogus precipitation data conform more exactly to the actual precipitation data. The general form of the equation used

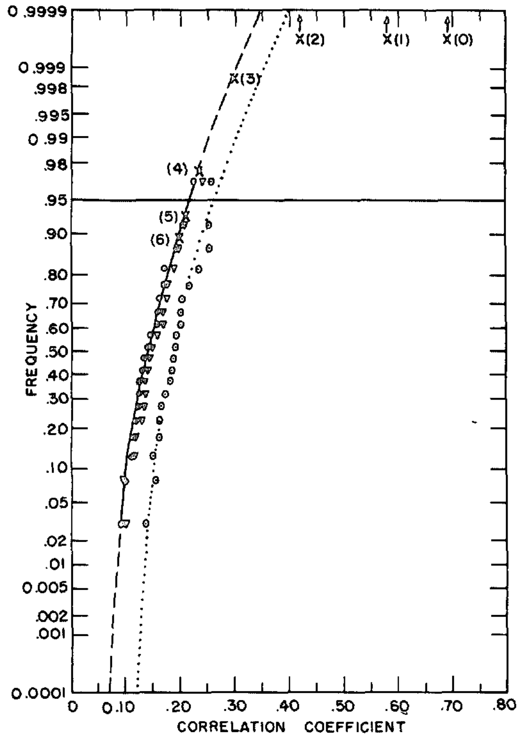


FIGURE 9.—Frequency distribution of the highest (in absolute value) of 499 correlations on each of 20 charts of correlations between 850-mb heights and bogus California precipitation values when the correlation between the precipitation values equals zero (dots), 0.13 (triangles), and 0.53 (open circles). The corresponding correlations between the same 850-mb heights and actually observed California precipitation on the same day, (0), and 1 through 6 days after the heights were observed, are shown by $X$s.

to transform random normal numbers into correlated random normal numbers, used by Gringorten (1968), is

$$y_0 = \eta_0 \qquad \text{for } i=0$$

and

$$y_i = \rho y_{i-1} + (1-\rho^2)^{1/2}\eta_i \qquad i \geq 1 \tag{5}$$

where $\eta_i$ is the $i$th normal number selected at random from the population without serial correlation, $\rho$ is the population correlation coefficient, and $y_0$ is equal to the initial random number.

In the development by Lund (1970) of the statistical significance test, it was assumed that the predictand values were not serially correlated. This was a reasonable assumption in the published example because the predictand values were spaced 1 yr apart. In this study, the Eastern United States precipitation values are not serially correlated but there is a day-to-day correlation of 0.53 and an average sample correlation of 0.13 for 4, 5, and 6 days after the initial day in the California precipitation values as shown in section 3.

The 8,020 random normal numbers mentioned above were transformed into two sets of correlated random normal numbers. One set had a serial correlation of 0.13 and the other had a serial correlation of 0.53. The correlated random normal numbers were processed one set at a time in exactly the same manner as the uncorrelated random normal numbers. The triangles plotted in figure 9 depict the highest correlation coefficients obtained using the bogus data with serial correlation of 0.13; the open circles depict corresponding values when the correlation was 0.53. The 1-day lag observed correlation, $X(1)$, should be compared with the dotted curve drawn through the open circles; the 2- and 3-day lags [$X(2)$ and $X(3)$, respectively] should be compared with curves drawn between the open circles and the triangles; and lags beyond 3 days should be compared with curves drawn between the triangles and the solid dots. Correlations exceeding the 0.0001 level of significance were found for periods through at least 2 days. The chosen significance level of 0.05 was exceeded for all lags of less than 5 days.

## 6. CONCLUDING REMARKS

Many models have been developed for the purpose of estimating the probability of precipitation occurrence as a function of the initial condition: precipitation or no precipitation. The Markov chain model described by Feyerherm and Bark (1967) and many others is the most popular. The probability of precipitation ($W$) at time $t + \Delta t$ given precipitation at time $t$, expressed $P(W_{t+\Delta t}|W_t)$, approaches 1 as $\Delta t$ approaches zero and approaches the climatic probability of precipitation $P(W)$ as $\Delta t$ becomes large. For small values of $\Delta t$, it has been shown by Jenkins and Twitchell (1959) and others, that the addition of more, or different, parameters other than $W_t$ into the equation does not significantly improve the probability estimates. It appears from the correlations between 850-mb heights ($Z$) and areal precipitation that there is some point in time, $\Delta t$ less than 24 hr, when $Z_t$ should be substituted for $W_t$ in the expression $P(W_{t+\Delta t}|W_t)$. Instead of replacing $W_t$ with $Z_t$, it may be better to add $Z_t$ and model $P(W_{t+\Delta t}|W_t,Z_t)$. The development of such a model is beyond the scope of this paper. The identification of locations of maximum correlations, shown in figure 4 for guidance in selecting points from which to consider 850-mb heights, is one of the limited objectives of this study.

It is shown that, if an accurate 850-mb prognostic chart can be prepared, approximately 50 percent (the square of the correlation coefficient 0.69) of the daily areal precipitation variability in the Sacramento and San Joaquin Valleys of California can be predicted using a simple statistical relationship with a single prognostic height. Without a prognostic chart, more than 33 percent of the variability can be predicted at least 1 day in advance with a single 850-mb height. Although a statistically significant amount of information is retained in the observed 850-mb heights for a period of at least 5 days, the few percent additional reduction in variance achieved is probably not operationally important.

The correlations between 850-mb heights and precipitation in the Eastern United States are lower than those obtained for California. With a prognostic chart, only about 19 percent of the variability can be predicted 1 day in advance; and without a prognostic chart, only about 10 percent can be predicted using just a single 850-mb height. The statistical significance of the correlation coefficients decreases to below the 5-percent level within 3 days.

Most, if not all, of the statistically significant predictors were found in areas which experienced synoptic weather forecasters would have suggested. The fact that the area of highest correlation between 850-mb heights and California precipitation remains essentially unchanged for several lags was somewhat surprising. This result may be partially due to the absence of data over the Pacific Ocean. There was no evidence that precipitation is foreshadowed by 850-mb heights observed at some unexpected distant location.

Only simple linear correlation coefficients between precipitation and 850-mb heights were included in this study, to simplify statistical significance testing and to identify the single predictor that might be selected to replace or supplement persistence in conditional climatologies of precipitation.

## ACKNOWLEDGMENTS

The expert computational support given by Eugene A. Bertoni Air Force Cambridge Research Laboratories, and Donald J. Armstrong, Analysis and Computer Systems, Inc., is gratefully acknowledged. This study would not have been possible without the grid-point data prepared and supplied by the U.S. Air Force Global Weather Central.

## REFERENCES

Blom, Gunnar, *Statistical Estimates and Transformed Beta-Variables*, John Wiley & Sons, Inc., New York, N.Y., **1958**, 176 pp.

Elliott, Robert D., "Extended-Range Forecasting by Weather Types," *Compendium of Meteorology*, American Meteorological Society, Boston, Mass., **1951**, pp. 834–840.

Feyerherm, A. M., and Bark, L. Dean, "Goodness of Fit of a Markov Chain Model for Sequences of Wet and Dry Days," *Journal of Applied Meteorology*, Vol. 6, No. 5, Oct. **1967**, pp. 770–773.

Gringorten, Irving I., "A Stochastic Model of the Frequency and Duration of Weather Events," *Journal of Applied Meteorology*, Vol. 5, No. 5, Oct. **1966**, pp. 606–624.

Gringorten, Irving I., "Estimating Finite-Time Maxima and Minima of a Stationary Gaussian Ornstein-Uhlenbeck Process by Monte Carlo Simulation," *Journal of the American Statistical Association*, Vol. 63, No. 324, Washington, D.C., Dec. **1968**, pp. 1517–1521.

Jenkins, Carl F., and Twitchell, Paul F., "Study and Evaluation of Two-Hour Objective Terminal Forecast Techniques," *Final Report*, Contract No. AF(604)-2450, Melpar, Inc., Falls Church, Va., Feb. **1959**, 219 pp.

Klein, William H., "An Objective Method of Predicting Quantitative Precipitation in the Tennessee and Cumberland Valleys," *Proceedings of the First National Conference on Statistical Meteorology, Hartford, Connecticut, May 27–29, 1968*, American Meteorological Society, Boston, Mass., May **1968**, pp. 20–28.

Klein, William H., Jorgensen, Donald L., and Korte, August F., "Relation Between Upper Air Lows and Winter Precipitation in the Western Plateau States," *Monthly Weather Review*, Vol. 96, No. 3, Mar. **1968**, pp. 162–168.

Krick, Irving Parkhurst, *Synoptic Weather Types of North America*, California Institute of Technology Press, Pasadena, **1943**, 237 pp.

Lund, Iver A., "Map-Pattern Classification by Statistical Methods," *Journal of Applied Meteorology*, Vol. 2, No. 1, Feb. **1963**, pp. 56–65.

Lund, Iver A., "A Monte Carlo Method for Testing the Statistical Significance of a Regression Equation," *Journal of Applied Meteorology*, Vol. 9, No. 3, June **1970**, pp. 330–332.

McAllister, C. R., "Cloud-Cover Recurrence and Diurnal Variation," *Journal of Applied Meteorology*, Vol. 8, No. 5, Oct. **1969**, pp. 769–777.

Russo, John A., Jr., Enger, Isadore, and Merriman, Guy T., "A Statistical Approach to the 12–48-Hr Prediction of Precipitation Probability," *Final Report* 7671-217, Contract No. Cwb 11100, The Travelers Research Center, Inc., Hartford, Conn., Aug. **1966**, 107 pp.

Snedecor, George W., and Cochran, William G., *Statistical Methods*, 6th Edition, The Iowa State University Press, Ames, **1967**, 593 pp. (see p. 185).

[*Received July 6, 1970; revised January 28, 1971*]